

LayoutGPT: Compositional Visual Planning and Generation with Large Language Models

Year: 2023 | Citations: 272 | Authors: Weixi Feng, Wanrong Zhu, Tsu-Jui Fu

Abstract

Attaining a high degree of user controllability in visual generation often requires intricate, fine-grained inputs like layouts. However, such inputs impose a substantial burden on users when compared to simple text inputs. To address the issue, we study how Large Language Models (LLMs) can serve as visual planners by generating layouts from text conditions, and thus collaborate with visual generative models. We propose LayoutGPT, a method to compose in-context visual demonstrations in style sheet language to enhance the visual planning skills of LLMs. LayoutGPT can generate plausible layouts in multiple domains, ranging from 2D images to 3D indoor scenes. LayoutGPT also shows superior performance in converting challenging language concepts like numerical and spatial relations to layout arrangements for faithful text-to-image generation. When combined with a downstream image generation model, LayoutGPT outperforms text-to-image models/systems by 20-40% and achieves comparable performance as human users in designing visual layouts for numerical and spatial correctness. Lastly, LayoutGPT achieves comparable performance to supervised methods in 3D indoor scene synthesis, demonstrating its effectiveness and potential in multiple visual domains.